THE UNIVERSITY OF CHICAGO

AN EMPIRICAL STUDY OF THE ABILITY TO GENERALIZE

A DISSERTATION SUBMITTED TO THE GRADUATE FACULTY
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

By
GEORGE MAXWELL PETERSON

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

Reprinted from
The Journal of General Psychology
Vol. VI, No. 1, 1932

R. UNIV.
BIBLIOTHEEK
LEIDEN

THE UNIVERSITY OF CHICAGO

AN EMPIRICAL STUDY OF THE ABILITY TO GENERALIZE

A DISSERTATION SUBMITTED TO THE GRADUATE FACULTY
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

By
GEORGE MAXWELL PETERSON

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

Reprinted from
THE JOURNAL OF GENERAL PSYCHOLOGY
Vol. VI, No. 1, 1932

Offprinted from *The Journal of General Psychology*, 1932, **6**, 90-114.

AN EMPIRICAL STUDY OF THE ABILITY TO GENERALIZE*

From the Department of Psychology of the University of Chicago

GEORGE M. PETERSON

INTRODUCTION

In this study we are concerned with a series of problems all involving a general principle which, if discovered and applied in their solution, will give the correct answer. The problems were given to primary and secondary school students with a view to discovering the trend of ability in solving them and the correlation of this ability with age, intelligence, and grade in school. We are also interested in any sex differences that may be present and in the incidence of the ability to state the general principle as we progress through the school grades.

The problems which we gave all involved the general principle of the law of the lever which may be stated: *Any lever is balanced when the moment on one side of the fulcrum is equal to the moment on the other side.* That is, a lever is balanced when the weight times the distance on one side is equal to the weight times the distance on the other.

APPARATUS

Our apparatus was a balance or lever of the first class, the arms of which were made of bakelite, each 9½ inches in length. Four hooks were placed at 2-inch intervals on the lower side of each arm. Running out consecutively from the fulcrum, and above each hook were the numbers 1, 2, 3, 4, painted in white. These were easily visible from the farthest distance in a normal-sized classroom.

Ten weights made of bakelite, one inch in diameter and weighing 1, 2, 3, 4, 5, 6, 8, 9, 10, and 12 ounces, respectively, were used. Thus the 12-ounce weight was about twelve times the length of the one-ounce weight. The appropriate numbers were painted in white on each weight.

*Recommended for publication by Harvey A. Carr, accepted by Carl Murchison of the Editorial Board, and received in the Editorial Office, September 3, 1931.

Information Concerning the Source of Data

Our data were all collected in the Chicago Public Schools. Since they do not have intelligence test records it was necessary for us to give our own test to obtain these. We used the Terman Group Test of Mental Ability for Grades 7 to 12, Form A. All testing was done in the afternoon. The intelligence test was given on the first afternoon in the case of each group and the weight problems were given on the second afternoon. Where two groups were obtained from the same school, which occurred in five instances, the intelligence tests were given successively on the same afternoon and the weight problems successively on the next afternoon. In no case did these groups in the primary schools have recess together, hence they could not communicate regarding the nature of the work. The only secondary school, Fenger High School, to contribute two groups had a five-minute passing period between the two test periods, but, since the groups were at the opposite ends of the building and neither knew of the other's participation in the work, we feel sure that the results were not affected by any communication between groups.

Table 1 gives a list of the schools, grades, and number of subjects used in this study.

TABLE 1

Primary Schools	Grade	Number of subjects
Cornell	5B	42
Bass	6B	42
Cornell	6A	43
Beale	7B	40
Burnside	7A	38
Bass	$8B_1$	35
Beale	$8B_2$	42
Burnside	8A	41
Secondary Schools	**Grade**	**Number of subjects**
Hyde Park	9B	39
Lindbloom	9A	32
Fenger	10B	31
Parker	10B	38
Fenger	10A	30
Tilden	11B	26
Englewood	11A	31
Bowen	12B	27

Our age and grade records were obtained from the intelligence test blanks which the students filled out. Age was computed in months. In a few instances the age did not check with the date of birth listed and in these cases the age record was always used, since it was considered that a child is more likely to know how old he is rather than in what year he was born.

Procedure

In giving the intelligence tests Terman's directions were followed.

In giving the weight problems the following procedure was followed:

After the apparatus was set up and papers passed for the students to write their answers on, the experimenter said:

> "Write your name on the paper you have just received. Notice that this paper has two columns on it, one headed 'problem number' and the other headed 'answer.' You are going to have twenty problems to solve. Be sure to get your answer in the right place.
>
> "All of these problems involve a general principle. That is, there is a method or way of doing them so that you can always get the problems right. It is up to you to discover what this way is as we go along. Then, after we have finished, I am going to ask you to write out this principle, that is, write out how you worked the problems. Now, I am going to show you a problem, then ask you to do one, then show you another, and then have you do another. It is when I show you that you must try to discover how to do them.
>
> "I have here a balance" (at this point I tipped it back and forth) "with hooks under each of these numbers as you go out from the center. These hooks are at equal distances as we go out. Now here is a set of weights" (here I held each one up in order). "Here is a 1-ounce weight with the number 1 on it, a 2-ounce weight with the number 2 on it, etc., up to a 12-ounce weight with the number 12 on it.
>
> "Now if I put an 8-ounce weight on hook 3" (I did this) "and wanted to balance it with a 12-ounce weight, I would put it on hook 2." (I did this.) "If I had put it on hook 1, it would not have been heavy enough." (I did this.) "If I had put it on hook 3 or 4 it would have been too heavy." (I did both.) "But on 2 it just balances." (I did this again.) "It is up to you to find out why this is, and when you know you can always get every problem right.

"Here is your first problem. If you had a 1-ounce weight on 4" (I put the 1 on 4), "where would you put a 2-ounce weight to balance it?" (I held up the 2 before them.) "Write your answer after problem No. 1 on the paper. That is, write down the number of the hook you think it should go on. Don't hurry. Take your time and figure carefully.

"Now I will show you another. If I had a 12-ounce weight on 1 and wanted to balance it with a 3-ounce weight, I would put it on 4." (I did this.) "Why is this? When you find out why you can get every problem right.

"Here is your second problem. If you had a 2-ounce weight on 3" (I put the 2 on 3), "where would you put a 6 to balance it?" (I held up the 6.) "Write your answer after problem No. 2."

This procedure continued until 20 problems were demonstrated and 20 problems were given to solve. In all, it took about 30 minutes to give the test, although no time limit was set. A new problem was not started until everyone had all the time he wanted on the old one.

The following order was given for the problems:

1. Demonstrated 8 on 3 12 on 2
   1. Given to solve 1 on 4 2 on (2)
2. Demonstrated 12 on 1 4 on 3
   2. Given to solve 2 on 3 6 on (1)
3. Demonstrated 12 on 3 9 on 4
   3. Given to solve 12 on 1 3 on (4)
4. Demonstrated 6 on 4 8 on 3
   4. Given to solve 4 on 3 6 on (2)
5. Demonstrated 9 on 2 6 on 3
   5. Given to solve 5 on 2 10 on (1)
6. Demonstrated 8 on 1 4 on 2
   6. Given to solve 8 on 1 2 on (4)
7. Demonstrated 3 on 4 6 on 2
   7. Given to solve 3 on 3 9 on (1)
8. Demonstrated 6 on 2 12 on 1
   8. Given to solve 4 on 4 8 on (2)
9. Demonstrated 6 on 1 3 on 2
   9. Given to solve 10 on 2 5 on (4)
10. Demonstrated 12 on 2 6 on 4
    10. Given to solve 2 on 2 4 on (1)
11. Demonstrated 4 on 1 2 on 2
    11. Given to solve 6 on 4 12 on (2)
12. Demonstrated 5 on 4 10 on 2
    12. Given to solve 3 on 2 6 on (1)
13. Demonstrated 8 on 2 4 on 4
    13. Given to solve 12 on 1 6 on (2)
14. Demonstrated 9 on 1 3 on 3
    14. Given to solve 6 on 2 3 on (4)
15. Demonstrated 2 on 4 8 on 1

15. Given to solve 6 on 3 9 on (2)
16. Demonstrated 10 on 1 5 on 2
16. Given to solve 4 on 2 8 on (1)
17. Demonstrated 6 on 2 4 on 3
17. Given to solve 8 on 3 6 on (4)
18. Demonstrated 3 on 4 12 on 1
18. Given to solve 9 on 4 12 on (3)
19. Demonstrated 6 on 1 2 on 3
19. Given to solve 4 on 3 12 on (1)
20. Demonstrated 2 on 2 1 on 4
20. Given to solve 12 on 2 8 on (3)

The numbers in parentheses indicate the answers which were credited as correct. It will be noticed that no problem is duplicated, and that the demonstrations are just the reverse of the problems given for solution, and given in reverse order. The students were never given the correct answers to the problems they had to solve. The only way they had of testing their principle was to apply it on the demonstrated problems to see if their answers agreed with those shown. In preliminary testing we tried the method of showing the correct answers to the problems after the students had written down their answers, and we found a number of cases where the answers had obviously been corrected.

After the 20th problem was given, I said: "Now turn over your sheet and write out the general principle, that is, write out how you did the problems."

The intelligence tests were graded according to Terman's directions. The weight problems were graded either right or wrong and the number of right answers was used as the individual's score. The statement of the principle was rated as either correct or incorrect. If the individual showed he had a practical working principle, full credit was given regardless of the abstractness of the statement. Thus statements such as "If I had an 8 on 3 I would put a 12 on 2 because 8 times 3 equals 12 times 2 or 24 equals 24," or "I multiplied the weight by the hook on one side and on the other to make them equal" or "I multiplied the weight by the hook on one side and divided by the other weight to get the number of the hook on the other side to put it on" were all given credit. Statements such as "I multiplied to get my answer" or "I divided" were not credited. No doubt, some of these had a working system since they received such high scores, but a number of others with low scores undoubtedly had no method well enough in mind to apply it successfully. In the case of a number of subjects from the Engle-

wood High School, where the group was a physics class, the statements were as abstract as those found in a physics book. The wonder is that 9 out of 31 in this class could not write a statement well enough to be credited, although most of these, no doubt, had a working system, as their high scores on the test indicate.

Summary of Related Literature

The literature in this general field is not very extensive. John C. Peterson (3) was the first, to our knowledge, to publish a study involving problems similar to ours, that is, problems involving a general principle which could be applied in their solution. He had a set of beads strung on a wire and required his subjects to draw alternately with him a given number of beads. The object was to have the last draw in reducing the number of beads to zero. The subjects could win if they drew to a multiple of the sum of the greatest and least possible draws. He had 46 subjects in all and gives an analysis of the method of solution, including introspective reports, in detail for one group of 14 of his subjects. His work relates to ours only incidentally where he makes age comparisons. His subjects in the group of 14 were adults with the exception of one 13-year-old high school boy. Only 2 of the 13 adults required fewer trials and problems than the boy whose time record was the lowest of the entire group.

Peterson says: "Subject iii (the boy) appeared to be less conscious of method, somewhat more reliant on trial and error procedure, and quicker in reaction than subjects i and ii; but in no respect did his responses vary so much from those of these two adults as did those of some of the other adult subjects."

In a preliminary experiment an 11-year-old boy in the seventh grade had a record equal to or superior to his father who also took part in the preliminary experiment. His record was also equal to that of the best subject in the group of 14, but the conditions were not quite the same.

These results might indicate that the ability to generalize does not increase with age but, in view of the fact that the comparisons are based on two cases, no confidence can be placed in the findings.

Joseph Peterson (4) tested the ability of 19 subjects in a problem requiring the association of letters and numbers. The correlation of rankings on "fewness of errors" to estimated intelligence was .86

and to marks based on six examinations in general psychology was .96.

In a later study (5) he gave the same type of problem to 81 college students. The correlation with psychology marks was .63±.04. This type of problem does not involve a general principle in the same sense that ours does. Mention is made of his studies merely because he regards the problems as involving rational learning and because of their correlation to intelligence and school marks.

Haught (2) continued Joseph Peterson's studies, using the same type of problem and also a modified form, and added two problems, a "checker puzzle" and the "Tait labyrinth puzzle." He intercorrelated these and also correlated each to Binet-Simon intelligence scores in the case of 74 college students. The intercorrelations ranged from .18 to .44, and the correlations with test intelligence ranged from .26 to .47. His conclusion pertinent for our study is that his rational learning problem and his modified rational learning problem test something not detected by intelligence tests but that the other two problems do not. He fails to give the reliability of any of his tests and we cannot estimate the influence of chance in determining the results.

Waters (7) used problems similar to those of John C. Peterson to study the influence of tuition on ideational learning. One of his groups received demonstrations as a form of guidance. In this case the experimenter drew first to win, giving two such demonstrations for each "set," after which the subjects drew first. This type of guidance was the only one to exert "a detrimental effect both on the subject's ability to master the first problem and also on his ability to deal with the second problem."

All of our groups were given demonstrations since it appeared to be the only way they could learn the principle under our conditions of group testing without giving them the principle directly. One group, however, was given only half as many demonstrations as the others and its mean score was significantly less than that of a comparable group. Waters did not vary the amount of demonstration to see if increasing it increased its efficacy.

Carroll (1) analyzed the errors made by fourth- and fifth-grade students in spelling unfamiliar words and concluded that bright students are able to generalize phonetically better than dull children. However, this type of generalization is not yet well defined since it cannot be stated as an invariable rule which will give a solution to

all the specific problems to which it might be applied and hence it cannot be measured directly.

Correlations

Pearson correlation coefficients were computed, using individual raw scores rather than grouping the data and using a correlation data sheet. All correlations not having a negative sign are positive. The following general formula was used:

$$r_{xy}=\frac{\frac{\Sigma_{xy}}{n}-\left(\frac{\Sigma_x}{n}\cdot\frac{\Sigma_y}{n}\right)}{\sigma_x\sigma_y}$$

Table 2 shows the value of the zero-order coefficients with their standard errors. They are carried to four places not because we consider these places significant but in order to secure closer agreement in our partial correlations dependent upon them.

TABLE 2
Zero-Order Coefficients

r and *P.E.*	Variables
r_{12} .5363±.031	age and intelligence
r_{13} .4587±.034	age and generalizing
r_{23} .4049±.036	intelligence and generalizing
r_{14} .8670±.011	age and grade
r_{24} .7573±.018	intelligence and grade
r_{34} .5584±.030	generalizing and grade

The subscripts of the correlation coefficients are as follows:

1 is age
2 is intelligence test scores
3 is weight problem scores
4 is grade in school

In general, the variable associated with the highest correlations is grade in school and the variable associated with the lowest correlations is weight problem scores.

Reliability

Since the correlations of the other variables with the weight problem scores are lower than we expected, indicating that we are apparently measuring something rather independent of these variables, and especially of age and intelligence test scores, we next

decided to find out if we were measuring anything at all, and hence tested the reliability of our measure. The only method at our disposal for this test was to correlate the odd and even items in the score. We then applied the Spearman-Brown formula to obtain the reliability for the original length of our test. The reliability was determined for each grade group and also for the entire range of data. The results are given in Table 3.

The average reliability coefficient for a single grade group is about .84. The average for the primary schools is .83 and for the secondary schools is .84. There is one instance of especially low reliability, that of the Englewood High School, where it is .50. This group was a physics class in which 22 out of 31 students received perfect scores, and must have equal odd and even scores. This fact is sufficient to explain the low correlation in this case and the coefficient is probably an underestimation of the reliability. If we do not include this low instance, the average reliability coefficient for all schools is .86 and the average for the secondary schools .90.

TABLE 3
RELIABILITY OF WEIGHT SCORES

Grade	r	By the Spearman-Brown Formula
5B	.64	.78
6B	.70	.82
6A	.82	.90
7B	.64	.78
7A	.59	.74
$8B_1$	.86	.92
$8B_2$	.71	.83
8A	.81	.89
9B	.87	.93
9A	.81	.89
10B	.86	.93
10A	.80	.89
11B	.67	.80
11A	.34	.50
12B	.92	.96
All grades	.907	.95

The reliability coefficient for the entire range of grades is .907, which becomes .95 by the Spearman-Brown formula. Although higher reliability would be desirable, we can be safe in assuming that our test is measuring something in which chance plays only a small part in determining the result or score.

Partial Correlations

We next computed the first-order partial coefficients necessary for those second-order partials in which we were interested. The first-order partials, with their standard errors, are given in Table 4.

It will be noticed that three of these coefficients become negative, those in which grade in school is partialed out. Among these are the correlations between weight problems and intelligence, namely, —.0615, and weight problems and age, namely, —.0332. Although these negative findings are too small to be significant, the reductions of the correlations to practically zero are interesting.

TABLE 4
First-Order Partial Coefficients

$r_{13.4}$ equals —.0615±.043	$r_{23.1}$ equals .2119±.041
$r_{12.4}$ equals —.3696±.037	$r_{34.1}$ equals .3619±.037
$r_{23.4}$ equals —.0332±.043	$r_{24.1}$ equals .6952±.022
$r_{13.2}$ equals .3130±.039	$r_{34.2}$ equals .4217±.035
$r_{14.2}$ equals .8361±.013	

Finally, we computed the second-order partial coefficients and their standard errors, shown in Table 4.

TABLE 5
Second-Order Partial Coefficients

$r_{13.42}$ equals —.0794±.043.	$r_{34.12}$ equals .3069±.039
$r_{13.24}$ equals —.0795±.043	$r_{34.21}$ equals .3071±.039
$r_{23.14}$ equals —.0603±.043	$r_{24.13}$ equals .6789
$r_{23.41}$ equals —.0603±.043	$r_{14.23}$ equals .8176

Two solutions are given as a check for the partial coefficients in which the weight problems are one of the unpartialed variables. We are mainly interested in these correlations, although the two coefficients, in which the weight problems are partialed out, are given for comparison.

It would appear from these findings that the variable we are most interested in, namely, the ability to solve problems involving a general principle, is not very closely related to any of the other three variables. We had not expected the correlation of this variable to intelligence test scores to become zero by partialing out the influence of grade in school and age. We had also expected age to be more closely related to this variable than is the case. When the other

variables are held constant, the correlation between grade in school and the weight problems is only .30, not a very high relationship. Reference to Table 7, comparing the standard deviations to the standard errors of estimate shows, in the case of this correlation, that the predictive value is low. The standard deviation of the weight problems for all subjects is 4.93 and the standard error of estimate is 4.69 when the correlation coefficient is .30.

Multiple Correlation and Prediction

In order to obtain the highest predictive value the three other variables had for the weight problems, we determined the multiple correlation coefficients. The results are shown in Table 6.

TABLE 6
Multiple Correlation Coefficients

$r_{3.12}$	equals .4958±.032
$r_{3.24}$	equals .5591±.030
$r_{3.14}$	equals .5607±.030
$r_{3.124}$	equals .5629±.029

Again, grade in school is shown to be the important variable in predicting scores on the weight problems. Intelligence adds little or nothing for predictive purposes, since, if we compare the standard errors of estimate to the standard deviation of the weight problems, given in Table 7, we see that the standard error of estimate for the combined age and grade variables is 4.08 in predicting weight problem scores, while adding the variable intelligence reduces this error to only 4.07.

Table 7 gives the value of all the correlations discussed so far, for the purpose of comparing the standard errors of estimate to the standard deviations. For those who are statistically minded, the comparisons may be made directly in terms of $k=\sqrt{1-r^2}$, the coefficient of alienation. We have also listed the value of the standard deviations and of the standard errors of estimate to illustrate how much is gained in predicting one variable when the value of the other is known.

The variable giving the least error for predicting scores in the weight problems is grade in school. Its error is, in fact, lower than that of the best combination of the other two variables, intelligence

TABLE 7
PREDICTIVE VALUE OF TESTS

Variables	r	k	σ of 1st variable in subscript of r	σ estimate 1st variable	σ of 2nd variable in subscript of r	σ estimate 2nd variable
r_{12}	.5363	.8440	23.44	19.78	39.43	33.28
r_{13}	.4587	.8885	23.44	20.83	4.93	4.38
r_{23}	.4049	.9143	39.43	36.05	4.93	4.51
r_{14}	.8670	.4983	23.44	11.68	3.94	1.96
r_{24}	.7573	.6530	39.43	25.75	3.94	2.57
r_{34}	.5584	.8295	4.93	4.09	3.94	3.27
$r_{13.4}$	—.0615	.9981	23.44	23.40	4.93	4.92
$r_{12.4}$	—.3696	.9292	23.44	21.78	39.43	36.64
$r_{23.4}$	—.0332	.9994	39.43	39.41	4.93	4.93
$r_{13.2}$	.3130	.9497	23.44	22.26	4.93	4.68
$r_{14.2}$	.8361	.5486	23.44	12.86	3.94	2.16
$r_{34.2}$	.4217	.9067	4.93	4.47	3.94	3.57
$r_{23.1}$	.2119	.9773	39.43	38.53	4.93	4.82
$r_{34.1}$	.3629	.9318	4.93	4.59	3.94	3.67
$r_{24.1}$	.6952	.7188	39.43	28.34	3.94	2.83
$r_{13.42}$	—.0794	.9968	23.44	23.36	4.93	4.91
$r_{13.24}$	—.0796	.9968	23.44	23.36	4.93	4.91
$r_{23.14}$	—.0603	.9982	39.43	39.36	4.93	4.92
$r_{23.41}$	—.0603	.9982	39.43	39.36	4.93	4.92
$r_{34.12}$	.3069	.9517	4.93	4.69	3.94	3.75
$r_{34.21}$	.3071	.9517	4.93	4.69	3.94	3.75
$r_{24.13}$	.6789	.7342	39.43	28.95	3.94	2.89
$r_{14.23}$	.8176	.5758	23.44	13.50	3.94	2.27
$r_{3.12}$	.4958	.8684	4.93	4.28		
$r_{3.24}$	.5591	.8291	4.93	4.09		
$r_{3.14}$	.5607	.8280	4.93	4.08		
$r_{3.124}$	.5629	.8265	4.93	4.07		

and age. The standard error in predicting weight problem scores from grade in school is 4.09, while the standard error of the multiple correlation involving intelligence and age is 4.28.

We want especially to call attention to the standard error of the multiple coefficient where all three variables are used to predict the weight problem scores. It is 4.07; the standard deviation for the

weight problems is 4.93. In other words, the multiple coefficient of .56 is not especially valuable for predicting scores on the weight problems.

Correcting for Attenuation

Knowing the reliability coefficient, we were able to correct our correlations for attenuation. We assumed that grade and age were measured perfectly reliably. Terman (6) reports a reliability of above .90 for the intelligence test for single grade ranges but does not report for our range. Since the weight problems have a reliability less than this for single grades, we assumed that the intelligence test would have at least the reliability of the weight problems for our range, and probably higher. We assumed a reliability of .95 for the lower limit and perfect reliability for the upper limit and found that these limits did not effect the corrected coefficients very markedly. The coefficients, corrected for attenuation, are given in Table 8.

TABLE 8

Correlations Corrected for Attenuation

r_{13}	equals	.47		
r_{34}	equals	.57		
r_{23}	equals	.41	(where R_{22} equals	1.00)
r_{23}	equals	.42	(where R_{22} equals	.95)

Very little is gained by these corrections. In fact, no matter how we manipulate our data statistically, we seem always to arrive at the conclusion that we are measuring something that is not measured to any great extent by our other variables.

Common sense would have us include the ability to derive and apply general principles as a part of the intellectual functions of human beings. We cannot prove that our problems measure this ability, but it seems logical to assume that they measure it rather than any other special ability. It may be that such ability is never general but always a function of the specific situations which involve different general principles. If we assume that we have at least measured one type of these specific situations, it appears, at least, that intelligence tests do not measure it. Apparently, in their attempt to measure "intelligence" they overlook one phase of it which would ordinarily be regarded as important.

Trends

In order to study trends of means, standard deviations, and coefficients of correlation, we have computed these measures separately for our different grade groups. The results are shown in Tables 9 to 13.

Table 9 gives the mean ages in months of our different grade groups, the standard errors of these means, and the standard deviations. The relation between mean age and grade in school is almost linear, with but one inversion at Grade 7A. Table 10 shows the same measures for the intelligence test scores. Here there are four inversions, at Grades 7B, 8A, 9A, and 10B. The same measures are shown for the weight problems in Table 11. Here there are five inversions, at Grades 7B, 7A, 10B, 10A, and 12B. There is a sixth at Grade 8A if we compare it to the Grade $8B_1$ group with a mean score of 13.00. However, the other $8B_2$ group has a mean score of 11.24, and it appears that the mean of 13.00 for Grade $8B_1$ is more likely to be in error than that of the 8A group.

In general, the means of the three variables increase as we progress through the school grades. None of the inversions are greater than three times the standard errors of the differences and may be assumed to be due to chance selection. There is no noticeable trend in the case of the standard deviations, these remaining about the same throughout the grades.

TABLE 9
Trend of Means of Age

Grade	*n*	Mean age	Standard error	Standard deviation
5B	42	132.1	1.87	12.11
6B	42	147.2	1.74	11.29
6A	43	147.9	1.67	10.98
7B	40	160.8	1.88	11.86
7A	38	158.0	1.81	11.13
$8B_1$	35	161.4	1.61	9.54
$8B_2$	42	164.7	2.31	14.98
8A	41	166.7	1.57	10.05
9B	39	168.9	1.67	10.45
9A	32	179.5	1.83	10.33
10B	31	184.7	2.24	12.50
10A	30	189.8	1.62	8.86
11B	26	202.6	1.86	9.49
11A	31	200.7	2.62	14.60
12B	27	204.9	1.82	9.48
All groups	539	168.3	1.01	23.44

TABLE 10
TREND OF MEANS OF INTELLIGENCE

Grade	*n*	Mean intelligence	Standard error	Standard deviation
5B	42	38.7	2.44	15.83
6B	42	62.9	3.81	24.72
6A	45	64.4	2.89	18.95
7B	40	63.6	3.42	21.66
7A	38	65.0	3.27	20.14
$8B_1$	35	86.8	4.39	25.97
$8B_2$	42	92.0	3.22	20.86
8A	41	88.6	4.36	27.90
9B	39	121.5	4.56	28.47
9A	32	117.6	3.96	22.41
10B	31	108.6	5.60	31.17
10A	30	126.7	5.70	31.21
11B	26	127.9	5.30	27.05
11A	31	130.9	4.92	27.38
12B	27	148.7	4.23	21.99
All groups	539	92.1	1.70	39.43

TABLE 11
TREND AND MEANS OF WEIGHT PROBLEM SCORES

Grade	*n*	Mean wt. problems	Standard error	Standard deviation
5B	42	9.50	.56	3.61
6B	42	9.76	.63	4.06
6A	43	10.88	.67	4.38
7B	40	10.23	.51	3.25
7A	38	9.97	.59	3.65
$8B_1$	35	13.00	.79	4.70
$8B_2$	42	11.24	.58	3.78
8A	41	12.12	.74	4.72
9B	39	14.05	.75	4.69
9A	32	15.72	.66	3.74
10B	31	14.03	.80	4.44
10A	30	15.50	.72	3.97
11B	26	17.92	.43	2.17
11A	31	19.26	.24	1.36
12B	27	17.04	.83	4.33
All groups	539	12.92	.21	4.93

Table 12 shows the correlation coefficients of age and intelligence (r_{12}), age and the weight problems (r_{13}), and intelligence and the weight problems (r_{23}). There standard errors are also given, showing that individually they do not possess much reliability. Taken collectively, however, they indicate certain tendencies. Thus, in the case of age and intelligence, the correlations are negative in 13 out

TABLE 12
TRENDS OF CORRELATION COEFFICIENTS

Grade	Age and intelligence r_{12}	Standard error	Age and wt. problems r_{13}	Standard error	Intelligence and wt. problems r_{23}	Standard error
5B	—.32	.14	—.12	.15	.15	.15
6B	—.52	.11	—.08	.15	.37	.13
6A	—.57	.10	—.25	.14	.53	.11
7B	—.30	.14	.37	.14	.20	.15
7A	—.37	.14	—.03	.16	.38	.14
$8B_1$	—.32	.15	—.04	.17	.26	.16
$8B_2$	—.55	.11	.04	.15	—.04	.15
8A	—.36	.14	—.04	.16	.30	.14
9B	—.51	.12	—.31	.14	.62	.10
9A	—.38	.15	—.003	.18	.04	.18
10B	—.39	.15	—.01	.18	.51	.13
10A	.03	.18	—.06	.18	.38	.16
11B	—.11	.19	—.05	.20	.17	.19
11A	—.29	.16	—.04	.18	.47	.14
12B	.13	.19	.04	.19	.09	.19

of 15 instances. Likewise, there is a negative relationship between age and the weight problems in 12 out of 15 cases. There is a positive relation between intelligence and the weight problems in 14 out of 15 cases.

If we partial out the influence of each of the other variables in their relation to the weight problems, the correlations become positive but low. Table 13 shows these results. The relation between age and weight problems with intelligence partialed out is positive in 12 out of 15 cases, and the relation between intelligence and the weight problems is positive in 14 out of 15 cases when age is partialed out.

It is interesting to note that these relations are positive when this direct method of holding the influence of grade constant is used, while the same relations are slightly negative when grade is partialed out by the indirect partial correlation technique. The discrepancies in the case of age and the weight problems are slight, but appear significant in the case of intelligence and the weight problems. Thus, the partial correlation of intelligence and the weight problems with age and grade partialed out was given at —.0795, or about —.08. The average of the 15 correlations, taken directly from the separate grades, between intelligence and the weight problems with age partialed out is about .31. We are inclined to place greater confidence in the latter since it is based on 15 instances, only one of which is negative.

Comparisons of Those Who Were Given Credit for the Principle to Those Who Were Not Given Credit

A comparison of the average scores made on the weight problems by those who were given credit for stating the principle to those who did not receive credit, is given in Table 14. The average score

TABLE 13
Trends of Partial Coefficients

Grade	$r_{13.2}$	$r_{23.1}$
5B	—.08	.12
6B	.14	.39
6A	.07	.49
7B	.46	.35
7A	.13	.42
$8B_1$	.05	.29
$8B_2$	.02	—.02
8A	.08	.31
9B	.09	.57
9A	.01	.04
10B	.24	.55
10A	—.05	.38
11B	—.03	.17
11A	.11	.48
12B	.03	.09

TABLE 14
Comparison of Means of Those Credited to Those Not Credited with Principle

Grade	Total mean	Mean of those credited	Mean of those not credited
5B	9.50	16.33	8.97
6B	9.76	15.75	9.13
6A	10.88	19.00	9.57
7B	10.23	13.80	9.71
7A	9.97	15.67	9.49
$8B_1$	13.00	17.08	10.87
$8B_2$	11.24	13.00	10.76
8A	12.12	17.50	9.90
9B	14.05	16.75	12.17
9A	15.72	17.65	12.50
10B	14.03	18.64	11.50
10A	15.50	18.07	12.93
11B	17.92	19.00	15.00
11A	19.26	19.68	18.22
12B	17.06	18.89	12.63
All groups	12.92	17.75	10.58
All primary schools	10.80	16.19	9.72
All secondary schools	16.08	18.44	13.02

of those who received credit was greater than that of those who did not receive credit in every case for the 15 separate grade groups. The standard errors were not calculated for these groups since the number of cases in the subgroups is so small as to make this concept useless. The reliability stands on the consistency of the reults.

The mean score of all those who received credit is 17.75, while that of all those who did not receive credit is 10.58. Of those who received credit, the mean score of those in the primary schools is 16.19, and of those in the secondary schools, it is 18.44. Also, of those who did not receive credit, the mean score among the primary-school students is 9.72 and, among the secondary-school students, it is 13.02. But the mean score of all primary-school students is 10.80 and, of all secondary-school students, it is 16.80. The larger increment in the last case than in the two preceding cases is due to the fact that more students received credit for the principle in the secondary schools than in the primary schools, a fact which is brought out in Table 15.

If we study the trends of means for both the credited and non-credited groups, we do not find as marked a differentiation as we get for the total groups. Furthermore, if we examine the inversions in the trend of the total means, namely, Grades 7B, 7A, 10B, 10A, and 12B, and compare, in Table 15, the percentage who received credit in these groups as against the percentage of their preceding groups, we find a decrease in every instance. Also, the high mean score of 13.00 for the $8B_1$ group is accompanied by a high percentage, namely, 34% of those credited with the principle.

Table 15 shows the incidence of those credited with the principle and the trend in the percentage of those credited as we go through the grade groups. It also shows the number of males and females in the groups and the number and percentage of each of these who were credited with the principle. There is an increase in the trend of percentages as we advance through the grades, but it is not very constant. The chance presence or absence of but a few individuals capable of grasping the principle of the problems can alter the value of the percentage markedly.

Sex Difference

Beyond the first two grade groups, the percentage of males credited with the principle is higher than females so credited in all but two instances, and in one of these the percentages are equal (see Table

TABLE 15
ANALYSIS OF THOSE CREDITED WITH THE PRINCIPLE

Grade	Total No.	No. males	No. females	Total No.	%	No. males	% males	No. females	% females
5B	42	24	18	3	7	1	4	2	11
6B	42	17	25	4	10	1	6	3	12
6A	43	23	20	6	14	4	17	2	10
7B	40	18	22	5	12	3	17	2	9
7A	38	18	20	3	8	3	17	0	0
$8B_1$	35	15	20	12	34	7	47	5	25
$8B_2$	42	20	22	9	21	6	30	3	14
8A	41	18	23	12	29	7	39	5	22
9B	39	16	23	16	41	6	38	10	43
9A	32	7	25	20	62	5	71	15	60
10B	31	22	9	11	35	9	41	2	22
10A	30	10	20	15	50	5	50	10	50
11B	26	26	0	19	73	19	73	—	—
11A	31	31	0	22	71	22	71	—	—
12B	27	4	23	19	70	4	100	15	65
All groups	539	269	270	176	33	102	38	74	27

TABLE 16
MEAN SCORES OF BOTH SEXES

Grade	Mean score: males	Mean score: females
5B	9.71	9.22
6B	10.23	9.44
6A	12.08	9.50
7B	10.56	9.95
7A	11.17	8.90
$8B_1$	14.27	12.05
$8B_2$	12.15	10.41
8A	13.44	11.09
9B	13.75	14.26
9A	16.00	15.64
10B	15.05	11.56
10A	16.30	15.10
12B	19.50	16.61

15). In the first two grades the incidence of those who were credited with the principle is so small that it has little weight in determining the scores made by the males and females. In the other cases the sex differences in these percentages correspond to the sex differences in scores on the weight problems.

Two of the groups, Grades 11B and 11A, were all males and cannot be used in studying sex differences. Table 16 shows the mean scores of the two sexes for the 13 remaining groups. Since the number

of individuals in each group is small, standard errors are not used. Greater reliance can be placed on the consistency of the result.

There is a difference in favor of the males in 12 of the 13 cases. If we assume the chances to be even that the males exceed the females, the chances of getting 12 such groups out of a total of 13 would be 13 out of 8192 or about 1 out of 630.

With the data which we have it was possible to test two hypotheses to account for these sex differences.

1. Since the correlations between age and the weight problems are slightly negative in all but three instances for the different grade groups (see Table 12), it might be that the females are slightly older except in these three instances and in the case where their weight score exceeds that of the males. Table 17 shows the differences in mean age expressed in terms of the males. A negative sign indicates that the females were older.

The females are older than the males in three instances, at Grades 7B, 7A, and 10B, in two of which they should be older to meet this hypothesis, namely, at 7A and 10B. The males are older in the other nine instances, although, according to the hypothesis, they should be older only at Grade 9B, where the sex differences favor the females, and in Grades $8B_2$ and 10A where the correlations are positive. Therefore, the facts fit the hypothesis in five out of 13 cases, at Grades 7A, $8B_2$, 9B, 10B, and 10A.

2. The second hypothesis is like the first with the substitution of intelligence for age. Since the correlations between intelligence and the weight problems are positive in all but one instance, at Grade $8B_2$ (see Table 12), if the sex differences are due to differences in intelligence, we should expect the males to have a higher intelligence score in all but this case and the case where the females have a higher weight problem score, at 9B. Table 17 shows that the males have a lower score at Grades 5B, $8B_1$, 9A, and 12B and a higher score at Grades $8B_2$ and 9B. Hence the facts fit this hypothesis in 7 out of 13 cases, at Grades 6B, 6A, 7B, 7A, 8A, 10B, and 10A.

The "age" hypothesis fails to account for sex differences at Grades 5B, 6B, 6A, 7B, $8B_1$, 8A, 9A, and 12B. The "intelligence" hypothesis fails at Grades 5B, $8B_1$, $8B_2$, 9B, 9A, and 12B. Thus, they both fail in four instances, at Grades 5B, $8B_1$, 9A, and 12B, and one tends to counteract the effect of the other at Grades 6B, 6A, 7B, 8A, $8B_2$, and 9A. In other words, the only grades in which both hypotheses

clearly succeed in accounting for the sex differences are 7A, 10B, and 10A; that is, in three cases out of thirteen.

On the other hand, if we examine the percentage of males and females who were credited with the principle (Table 15), we find that a higher percentage of females received credit only at Grades 5B, 6B, and 9B. The percentages are equal at Grade 10A. We have already shown that those who receive credit tend to make higher scores than those who do not receive credit (Table 14). Since the females have a higher mean score at Grade 9B than the males, the sex differences in mean scores coincide with sex differences in percentages credited with the principle in all but two cases, at Grades 5B and 6B. In these two cases the number receiving credit is so small that the chance presence or absence of one individual would markedly alter the percentage. We seem justified in concluding, then, that the sex differences in favor of the males are significant though slight, and cannot be accounted for by differences in intelligence or age.

The Influence of Tuition

In our introduction we mentioned that we had one group which was given only 10 demonstration problems. This group has not been included in any of the foregoing work. It was tested at Parker High School, Grade 10B. Table 18 compares the results of this group to the 10B group which had 20 demonstration problems. The mean age and intelligence scores of Parker were higher than

TABLE 17
Sex Differences in Favor of Males

Grade	Diff. in wt. problem	Age diff.	Intelligence diff.	Diff. in % credited
5B	.49	5.56	— 3.02	— 7
6B	.79	.71	13.19	— 6
6A	2.58	2.64	8.64	7
7B	.61	— 1.81	8.40	8
7A	2.27	— 3.43	9.39	17
$8B_1$	2.22	6.95	— 8.22	22
$8B_2$	1.74	5.89	2.39	16
8A	2.35	5.81	5.74	17
9B	— .51	5.43	.81	— 5
9A	.36	2.79	— 3.91	11
10B	3.49	— 4.44	10.53	19
10A	1.20	1.55	20.70	0
12B	2.89	10.40	—45.48	35

TABLE 18
INFLUENCE OF TUITION

	Mean age	Mean intell.	Mean wt. prob.
Parker 10B (10 demonstrations)	189.26	121.08	12.76±.69
Fenger 10B (20 demonstrations)	184.70	108.70	14.03±.80
Difference			1.27
Standard error of difference			1.05
Ratio or σ value of difference			1.21
Area of curve below 1.21 σ			.11314

those of the normal group, so that we should also expect the mean score on the weight problems to be somewhat higher if the number of demonstrations given had no influence on the score. However, this score is lower by 1.27, whereas the standard error of the difference is 1.05. It is, therefore, 1.21 times as large as the standard error. The area of the normal curve below 1.21 σ is .11314. We may then say that the chances are about 1 out of 9 that the larger mean score of our normal group is due to chance. The only factor that we know to have varied and which we can assign this difference to is the greater amount of demonstrational guidance of the normal group.

SUMMARY AND CONCLUSIONS

1. The ability to solve problems involving a general principle may be measured with considerable reliability, indicating that we are measuring a variable in which chance plays only a small part in determining the score of an individual.

2. There is an interrelation which results in positive coefficients of correlation between the four variables measured in this study, namely, age, intelligence test scores, the ability to solve problems involving a general principle, and grade in school.

3. There is no relation between age and the ability to solve problems involving a general principle when grade in school is held constant, either indirectly by means of the partial correlation technique, or directly by computing the correlations between age and the weight problem scores for each grade.

4. When grade in school is held constant indirectly by means of the partial correlation technique, there is no relation between intelligence scores and the ability to solve problems involving a

general principle. But when grade in school is held constant directly by computing the correlations between intelligence test scores and the weight problem scores for each grade, some positive relation is found between the two variables. However, it is still not high enough to indicate very much in common between these two tests. We therefore seem to be measuring a trait commonly thought of as intellectual but which has been neglected by ordinary intelligence tests.

5. There is some positive relation, but not a great deal, between grade in school and the ability to solve problems involving a general principle when age and intelligence are both partialed out. Grade in school also gives the highest zero order coefficient with the weight problem scores. It appears, then, that schooling is more closely related than intelligence, as the latter is measured by the usual type of test, to an ability which many would regard as a significant intellectual function. Our measure of this function involves a principle which is considered important both scientifically and practically by the physicist, engineer, and educator. The schools seem to give a type of training which is general in its effect upon this specific ability, since only one of our groups (Englewood 11A) had specific training involving this physical principle.

6. None of the correlations are very valuable for predicting scores on the weight problems. The best combination of the other three variables has an alienation coefficient of .8265.

7. There is an increase in mean scores in the weight problems as we progress through the school grades. There is also a correlated increase in the percentage of those who understood the principle. These two variables are not perfectly associated, however, since the mean scores of those who were not credited with the principle also increases as we progress through the school grades.

8. There are slight sex differences favoring the males in the ability to solve problems involving a general principle. These differences cannot be accounted for in terms of age or intelligence differences.

9. When the amount of demonstrational guidance is decreased, there is a decrease in the ability to solve problems involving a general principle.

References

1. Carroll, H. A. Generalization of bright and dull children, a comparative study with special reference to spelling. *J. Educ. Psychol.*, 1930, **21**, 489-499.
2. Haught, B. F. The interrelation of some higher learning processes. *Psychol. Monog.*, 1922, **30**, No. 139. Pp. 71.
3. Peterson, J. C. The higher mental processes in learning. *Psychol. Monog.*, 1920, **28**, No. 129. Pp. 121.
4. Peterson, J. Experiments in rational learning. *Psychol. Rev.*, 1918, **25**, 441-467.
5. ———. The rational learning test applied to eighty-one college students. *J. Educ. Psychol.*, 1920, **11**, 137-150.
6. Terman, L. M., Burks, B. S., & Jensen, D. W. Genetic studies of genius. Vol. 3. The promise of youth. Stanford University, Calif.: Stanford Univ. Press, 1930. Pp. xiv+508.
7. Waters, R. H. The influence of tuition upon ideational learning. *J. Gen. Psychol.*, 1928, **1**, 534-549.

University of Denver
Denver, Colorado

UNE ÉTUDE EMPIRIQUE DE LA CAPACITÉ À GÉNÉRALISER

(Résumé)

On a fait subir une série de vingt problèmes employant tous le principe de la loi du levier à 577 élèves de l'école élémentaire des "grades" 5 à 12. Alternativement avec ces problèmes on a fait subir vingt problèmes de démonstration pour le rendre possible pour les sujets de dériver le principe et de l'appliquer aux problèmes à résoudre. On a fait une étude de la tendance de la capacité des élèves des divers "grades" de les résoudre, de la corrélation de cette capacité avec les résultats des tests d'intelligence, l'âge, et le "grade" scolaire, de la croissance dans l'incidence de la capacité de donner la principe, des différences de sexe, et au cas d'un groupe d'un "grade," de l'influence de décroître le quantité de direction démonstrative. On a trouvé que (a) la capacité de résoudre les problèmes employant un principe général peut être mesurée d'une façon assez constante, (b) il y a une intercorrélation qui donne des coéfficients positifs de corrélation entre l'âge, le "grade," les résultats des tests d'intelligence, et la capacité de résoudre ces problèmes, mais ces coéfficients sont réduits, en quelques cas pratiquement à nul, par la technique de corrélation partielle, et aucuns ne possèdent une valeur pour les buts prédictifs, (c) il y a une croissance dans les résultats moyens des problèmes et dans le nombre de sujets qui ont compris le principe avec l'avancement dans le "grade" scolaire, (d) il y a de petites différences de sexe qui favorisent les mâles, (e) une décroissance des démonstrations décroît le résultat moyen des problèmes.

Peterson

EINE EMPIRISCHE UNTERSUCHUNG DER FÄHIGKEIT DES GENERALISIERENS

(Referat)

Es wurden 577 Schülern der fünften bis zwölften Klassen der öffentlichen Schulen [d.h. 9 bis 19-jährigen Schülern] eine Serie von 20 Aufgaben vorgelegt die alle das Prinzip des Hebelgesetzes [law of the lever] in Anspruch nahmen. Es alternierten mit diesen Aufgaben 20 Demonstrations aufgaben die gegeben wurden, um die Vpp. dazu zu befähigen, den Grundsatz zu ermitteln und ihn auf die zu lösenden Aufgaben anzuwenden. Man untersuchte die Richtung der Variationen in der Fähigkeit der Schüler der verschiedenen Klassen die Aufgaben zu lösen, die Korrelation dieser Fähigkeit mit an Intelligenzprüfungen erhaltenen Zahlen und mit Alter und Schulstadium, die Zunahme der Verbreitung der Fähigkeit, den Grundsatz zu formulieren, Geschlechtsunterschiede, und, bei einer der Klassengruppen, die Einwirkung der Verminderung der Quantität der Lenkung durch Beweismaterial. Die Befunde waren folgende: (*a*) die Fähigkeit, Aufgaben zu lösen die einen Grundsatz in Anspruch nehmen lässt sich ziemlich zuverlässig messen. (*b*) Es gibt eine Interkorrelation die positive Koeffiziente verursacht zwischen Alter, Klasse, Zahlen an Intelligenzprüfungen, und Fähigkeit diese Aufgaben zu lösen. Diese Koeffizienten werden aber herabgesetzt, in manchen Fällen fast bis auf Nul, durch das Verfahren der Teilkorrelation (partial correlation) und sie sind in keinem Falle zur Vorhersagung nützlich. (*c*) Es zeigt sich mit Erhöhung der Schulklasse eine Erhöhung der an den Aufgaben erzielten mittleren Zahlen und der Zahl der den Grundsatz verstehenden Versuchspersonen. (*d*) Es zeigen sich geringe Geschlechtsunterschiede zu Günsten der männlichen Vpp. (*e*) Eine Verminderung der Beweisführungen vermindert die an den Aufgaben erhaltenen mittleren Zahlen.

PETERSON